

Pathology-Informed Latent Diffusion Model for Anomaly Detection in Lymph Node Metastasis

Jiamu Wang¹, Keunho Byeon¹, Jinsol Song¹, Anh Nguyen¹, Sangjeong Ahn²,
Sung Hak Lee^{3*}, and Jin Tae Kwak^{1*}

¹ School of Electrical Engineering, Korea University, Seoul 02841, Korea
{taurusmumu, bkh5922, truetg, ngtienanh, jkwak}@korea.ac.kr

² Department of Pathology, Korea University Anam Hospital and Department of Biomedical Informatics, Korea University College of Medicine, Seoul 02841, Korea
vanitas80@korea.ac.kr

³ Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
hakjjang@catholic.ac.kr

Abstract. Anomaly detection is an emerging approach in digital pathology for its ability to efficiently and effectively utilize data for disease diagnosis. While supervised learning approaches deliver high accuracy, they rely on extensively annotated datasets, suffering from data scarcity in digital pathology. Unsupervised anomaly detection, however, offers a viable alternative by identifying deviations from normal tissue distributions without requiring exhaustive annotations. Recently, denoising diffusion probabilistic models have gained popularity in unsupervised anomaly detection, achieving promising performance in both natural and medical imaging datasets. Building on this, we incorporate a vision-language model with a diffusion model for unsupervised anomaly detection in digital pathology, utilizing histopathology prompts during reconstruction. Our approach employs a set of pathology-related keywords associated with normal tissues to guide the reconstruction process, facilitating the differentiation between normal and abnormal tissues. To evaluate the effectiveness of the proposed method, we conduct experiments on a gastric lymph node dataset from a local hospital and assess its generalization ability under domain shift using a public breast lymph node dataset. The experimental results highlight the potential of the proposed method for unsupervised anomaly detection across various organs in digital pathology. Code: <https://github.com/QuIIL/AnoPILaD>.

Keywords: Unsupervised Anomaly Detection · Diffusion Model · Visual-Language Model · Lymph Node Metastasis.

1 Introduction

Lymph node metastasis is a crucial prognostic factor in cancer progression and treatment decisions [1]. With the advent of digital pathology, several artificial

* These authors contributed equally to this work

intelligence approaches have been proposed to automate the detection of lymph node metastasis within tissues. Many of these methods are based on supervised learning methods, leveraging convolutional neural networks (CNNs) and transformer-based models to identify metastasis with high accuracy [2]. While effective, these methods heavily depend on exhaustive expert annotations, which are time-consuming and resource-intensive. To address this limitation, unsupervised learning has emerged as a viable alternative, as it does not demand manual annotations. In unsupervised learning, a model is trained solely on normal (in-distribution) samples to learn a representation of in-distribution patterns. The trained model then detects abnormal or out-of-distribution (OOD) samples by identifying deviations from the learned in-distribution patterns, making it particularly well-suited for large-scale applications in digital pathology, where annotated abnormal samples are scarce [3].

Recently, generative models have been widely used for unsupervised anomaly detection, with two popular approaches: density-based methods and reconstruction-based methods. Density-based methods, such as variational autoencoders (VAEs) [4–7], learn representations of in-distribution data, assigning higher likelihoods to in-distribution samples and lower likelihoods to OOD samples. In contrast, reconstruction-based methods are trained exclusively on normal data to guarantee poor reconstruction quality for abnormal samples and high reconstruction quality for normal samples. These include autoencoder (AE)-based models [8–10], Generative Adversarial Network (GAN) [11–13]-based models and denoising diffusion probabilistic model (DDPM)-based [14–17] models.

In this paper, we introduce AnoPILaD, a Pathology-Informed Latent Diffusion model for anomaly detection in lymph node pathology images. This framework combines a latent diffusion model (LDM) [18] and a vision-language model (VLM) [19] for an improved identification of anomalies in pathology images. AnoPILaD utilizes a LDM to learn a compact representation of normal images in a latent space via iterative diffusion and denoising processes while preserving critical histopathological features. AnoPILaD also adopts a VLM to select pathology-specific normal keywords, semantically guiding the reconstruction process towards a specific direction. In this manner, AnoPILaD achieves small deviations for normal samples and large deviations for abnormal samples, enhancing the accuracy and robustness of anomaly detection.

2 Methods

2.1 Pathology-Informed Latent Diffusion Model

Recent research often employs DDPMs for reconstruction-based anomaly detection. A trained DDPM p_θ generates samples that match the in-distribution patterns $\mathbf{z}_0 \sim q(\mathbf{z}_0)$ by adding noise in a forward process (diffusion process), which has tractable posteriors at time $t - 1$:

$$q(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{z}_0) = \mathcal{N}\left(\mathbf{z}_{t-1} \middle| \tilde{\mu}(\mathbf{z}_t(\mathbf{z}_0, \epsilon), t), \tilde{\beta}_t \mathbf{I}\right) \quad (1)$$

where $t \sim [1, 1000]$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, \mathbf{z}_t is a noisy sample, and $\tilde{\beta}_t$ is a predefined constant. The model progressively removes noise in a reverse process:

$$p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t) = \mathcal{N}\left(\mathbf{z}_{t-1} \middle| \mu_\theta(\mathbf{z}_t(\mathbf{z}_0, \epsilon_\theta), t), \tilde{\beta}_t \mathbf{I}\right) \quad (2)$$

where ϵ_θ is an approximator to predict ϵ from \mathbf{z}_t . The model is trained by decreasing the KL divergence between two Gaussians, and a simple objective function is given by:

$$\mathcal{L} = E_{\mathbf{z}_t, t, \epsilon \sim \mathcal{N}(0, \mathbf{I})} \left[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, t)\|_2^2 \right]. \quad (3)$$

A previous work [17] adopted a DDPM for detecting breast lymph node metastasis, referred to as AnoDDPM, where normal samples are in-distribution data and metastasis samples are OOD data. The pretrained DDPM denoised partially diffused inputs $\mathbf{z}_t \sim q(\mathbf{z}_t|\mathbf{z}_0)$ while steering the reverse process toward in-distribution patterns and get the reconstruction $\hat{\mathbf{z}}_0 \sim p_\theta(\mathbf{z}_0|\mathbf{z}_{1:t})$. Then, the discrepancy between the input \mathbf{z}_0 and its reconstruction $\hat{\mathbf{z}}_0$ serves as the anomaly score, which is expected to be small for normal samples and large for metastasis samples. Though successful, AnoDDPM exhibited a substantial number of false positive [17], indicating its insufficient ability to differentiate samples from normal and OOD distributions (Fig. 1). To address this issue and improve anomaly detection performance, AnoPILaD integrates a LDM with pathology-specific textual prompts, based on the assumption that these prompts enhance reconstruction quality and, in turn, magnify the contrast between normal and abnormal samples. We train a LDM with the following objective:

$$\mathcal{L} = E_{\mathbf{z}_t, t, c, \epsilon \sim \mathcal{N}(0, \mathbf{I})} \left[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, t, c)\|_2^2 \right] \quad (4)$$

where c denotes a textual prompt. We conduct pathology-informed reconstruction for lymph node pathology images by using the same reconstruction procedure with AnoDDPM. Only exception is that the reverse process is guided by the textual condition, which is given by $\hat{\mathbf{z}}_0 \sim p_\theta(\mathbf{z}_0|\mathbf{z}_{1:t}, c)$.

2.2 Weighted Prompts Generation

To introduce a stronger inductive bias to the reconstruction process, we propose to exploit prior pathology knowledge of normal lymph nodes. Specifically, we collect a pool of 74 pathology keywords from the literature, describing characteristics of cells and microenvironments of normal lymph nodes and tissues. These keywords are reviewed and validated by an experienced pathologist to ensure clinical relevance and accuracy.

Given a pathology image, we align it with the pathology keywords, identify the most relevant keywords, and use them to generate a text prompt, guiding the reconstruction process towards a pathology-informed direction (Fig. 2). To achieve the alignment between pathology images and keywords, we adopt

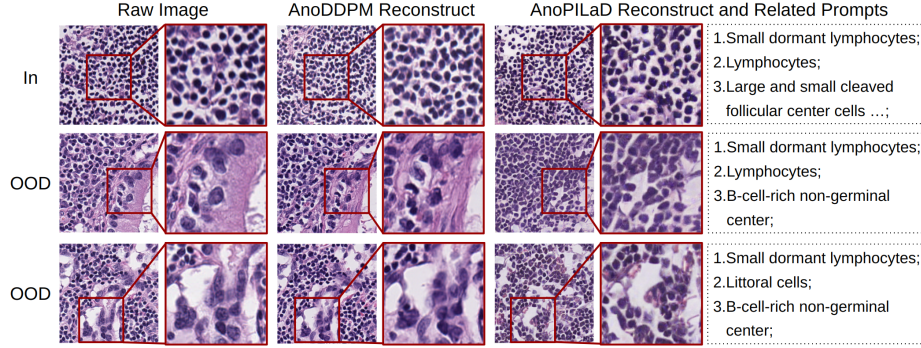


Fig. 1. Lymph node image reconstruction using diffusion-based methods. Using text prompts, AnoPILaD produces normal-like pathology images for both in-distribution (in) and out-of-distribution (OOD) samples.

CONCH [20], a vision-language foundation model pre-trained on over 1.17 million pathology image-caption pairs. The image encoder and text encoder of CONCH are used to produce image and text embeddings for each pair of input image and keywords, respectively. Then, we compute cosine similarity scores between image embeddings and text embeddings and choose top-five most similar keywords. The similarity scores of the selected keywords are subsequently normalized by dividing by the median score. Using the selected keywords and their normalized scores, we generate a weighted prompt as shown in Fig. 2. The weighted prompt is transformed into an embedding vector, which is fed into the LDM, following the procedure illustrated in the Compel library ^{*}.

Table 1. Label Distribution of Two Datasets

Dataset	WSI-level dataset					Patch-level dataset				
	D_{tr}^w	D_{val}^w	D_{in}^w	D_{out}^w	$(D_{out,m}^w)$	D_{tr}^p	D_{val}^p	D_{in}^p	D_{out}^p	$(D_{out,m}^p)$
LH	643	50	58	57		1,373,475	102,240	174,703	115,330	
C16	-	32	80	49	(22)	-	37,056	240,139	55,659	(55,536)

3 Experiment

3.1 Dataset

We obtained a gastric lymph node dataset from two local hospital (LH) sites containing 808 Whole Slide Images (WSIs), where 751 of them are normal and 57 are metastasis WSIs with partial annotation. All the metastasis WSIs are

^{*} <https://github.com/damian0815/compel>

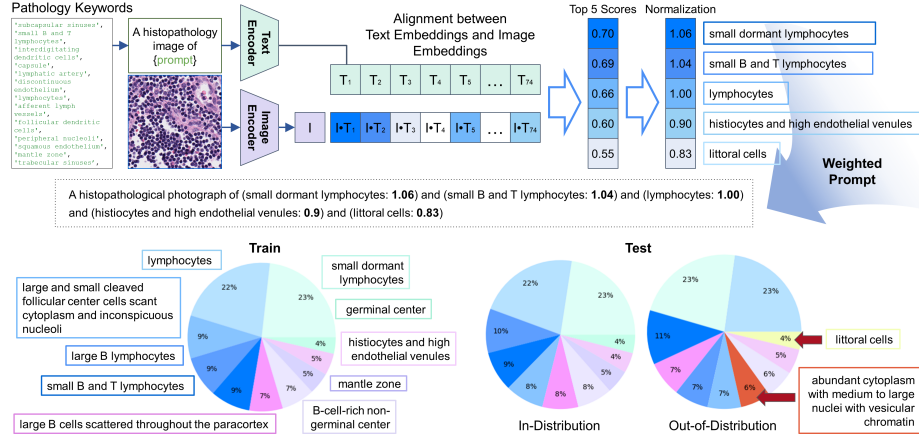


Fig. 2. (Top) Illustration of generating weighted text prompts. (Bottom) Distribution of top-10 frequent pathology keywords in the training and test sets the local hospital dataset.

used as an OOD test set ($D_{out}^{LH,w}$). The normal WSIs are split into a training set of 643 WSIs ($D_{tr}^{LH,w}$), a valid set of 50 WSIs ($D_{val}^{LH,w}$), and an in-distribution test set of 58 WSIs ($D_{in}^{LH,w}$). We further divide WSIs into patches, resulting in $D_{tr}^{LH,p}$ (1,373,475), $D_{val}^{LH,p}$ (102,240), $D_{in}^{LH,p}$ (174,703 patches, 138054 from $D_{tr}^{LH,w}$ and 36649 from normal annotated $D_{out}^{LH,w}$), and $D_{out}^{LH,p}$ (115,330 with fully metastasis annotation). Moreover, we employed the Camelyon16 Challenge dataset (C16) [21], a breast lymph node dataset, for independent testing. We utilize 32 normal WSIs as a valid set ($D_{val}^{C16,w}$), 80 normal WSIs as an in-distribution test set ($D_{in}^{C16,w}$), and 49 metastasis WSIs as an OOD test set ($D_{out}^{C16,w}$). Among the 49, 22 WSIs contain tumor cell cluster larger than 2 mm in diameter, designated as C16 Macro ($D_{out,m}^{C16,w}$). All tumor regions in LH are larger than 2 mm in diameter. We further divide WSIs into patches, resulting in $D_{val}^{C16,p}$ (37,056), $D_{in}^{C16,p}$ (240,139), and $D_{out}^{C16,p}$ (55,659 patches with fully metastasis annotation). The base prevalence are 0.3976 ($D_{in}^{LH,p}$); 0.4956 ($D_{tr}^{LH,w}$); 0.1882 ($D_{in}^{C16,p}$); 0.3798 ($D_{in}^{C16,w}$); 0.2157 ($D_{out,m}^{C16,w}$). For patch-level evaluation, WSIs from both LH and C16 are divided into 256×256 pixel patches using the pixel-level annotations. All WSIs are processed at $20 \times$ magnification. An overview of the datasets is presented in Table 1.

3.2 Experiment Details

To implement AnoPILaD, we utilized the stable diffusion model v1.5. The diffusion model was fine-tuned with Adam optimizer, a learning rate of $1e-5$, and a batch size of 64 using low-rank adaptation (LORA) with the update matrices dimension of 4. We used FID (Frechet Inception Distance) [22] to measure the

quality of the images generated by the diffusion model. At 400k steps, the model achieved the lowest FID, which was chosen for subsequent testing. The input image size was 256×256 pixels. We compared AnoPILaD with both density- and reconstruction-based OOD methods. For density-based methods, we used the negative likelihood (NLL) of a VAE backbone [4] and its three variants: Regret [5], LLR [6], and complexity [7]. The latent vector size was set to 100, and 64×64 pixels input images were randomly cropped. As for reconstruction-based methods, we employed f-AnoGAN [13], AE, and MemAE [10]. f-AnoGAN used randomly cropped 64×64 input image patches. For AE and MemAE, they follow the architectural design of [10] and input image size was 256×256 pixels. Both AnoDDPM and AnoPILaD were implemented using Diffusers library [23] and used a PLMS sampler [24] with 100 timesteps in inference. Unless otherwise specified, the training procedures followed those outlined in the original work.

For each method, we calculated z-scores of the anomaly scores for all the patches in each testing set and evaluated patch-level OOD detection by computing area under a receiver operating characteristic curve (AUC) and area under the precision-recall curve (AUPR). For WSI-level evaluation, we produced a z-score heatmap and applied a morphological erosion operation [25] with a 2×2 window since metastasis area can be very small. We then assessed the classification and segmentation performance of each model. For classification, we adopted two methods: maximum z-score (Z_{MAX}) and average of 99th percentile z-scores (Z_{99}). These are used to calculate WSI-level AUC and AUPR. For segmentation, we calculated the mean patch-level DICE and intersection-over-union (IoU) of D_{out}^w to assess the overlap between the predictions and annotations, and the mean patch-level true negative rate (TNR) of D_{in}^w since there is no positive regions. The segmentation prediction threshold was zero. To decide reconstruction timestep for two diffusion model-based methods, we tested eight timestep values as in [17] and chose the timestep according to the best WSI classification performance on LH test set, which is 674 for both methods.

4 Results

We trained AnoPILaD and all competing models with only D_{tr} and separately evaluated their patch-level and WSI-level performance on two distinct datasets originated from different organs. Thus, this evaluation provides insights into the robustness of the models against domain shifts due to variations in tissue types.

Table 2. Patch-level Anomaly Detection Results

		NLL	Regret	LLR	complexity	f-AnoGAN	AE	MemAE	AnoDDPM	AnoPILaD
LH	AUC	0.4982	0.6720	0.6078	0.7931	0.2289	0.9254	0.9290	0.8555	0.9587
	AUPR	0.5552	0.6718	0.6260	0.7139	0.3377	0.8906	0.8886	0.7841	0.9499
C16	AUC	0.3250	0.6480	0.7065	0.7752	0.1735	0.6584	0.6611	0.6857	0.8884
	AUPR	0.1600	0.3441	0.4765	0.5140	0.1104	0.4759	0.4880	0.5741	0.6987

Table 2 presents the performance of patch-level anomaly detection. AnoPILaD substantially outperformed all other methods. Among the four density-based methods (NLL, Regret, LLR, and complexity), NLL failed to distinguish OOD patches from normal patches. Its variants improved performance, with the complexity method achieving the largest improvement. Despite these improvements, complexity remained substantially inferior to AnoPILaD, with a large gap of ~ 0.16 AUC and ~ 0.23 AUPR. Among the reconstruction-based methods, f-AnoGAN demonstrated the poorest performance, while AE-based methods (AE and MemAE) achieved the highest scores. In LH, their AUC and AUPR scores were approximately 0.03 and 0.05 lower than AnoPILaD, respectively. However, in C16, the performance gap substantially increased to ~ 0.22 in AUC and ~ 0.22 in AUPR, indicating the superior robustness of AnoPILaD against domain shifts due to differences in organ types.

Table 3. WSI-level Anomaly Detection Results

		LH			C16			C16 Macro		
<i>Classification</i>		AUC	AUPR	-	AUC	AUPR	-	AUC	AUPR	-
AE	Z_{max}	0.9622	0.9612	-	0.6612	0.5961	-	0.6398	0.3677	-
	Z_{99}	0.9395	0.9381	-	0.5798	0.5217	-	0.5523	0.2918	-
MemAE	Z_{max}	0.9504	0.9440	-	0.6505	0.5689	-	0.6381	0.3657	-
	Z_{99}	0.9365	0.9382	-	0.5686	0.5313	-	0.5597	0.3141	-
AnoDDPM	Z_{max}	0.7840	0.6616	-	0.4992	0.3885	-	0.4926	0.2146	-
	Z_{99}	0.9383	0.8995	-	0.4551	0.3905	-	0.5119	0.2347	-
AnoPILaD	Z_{max}	0.9837	0.9740	-	0.6745	0.6140	-	0.8062	0.5965	-
	Z_{99}	0.9943	0.9948	-	0.6367	0.5902	-	0.8023	0.5886	-
		LH			C16			C16 Macro		
<i>Segmentation</i>		TNR	DICE	IoU	TNR	DICE	IoU	TNR	DICE	IoU
AE		0.7981	0.3812	0.2902	0.4536	0.1745	0.1317	0.4536	0.3249	0.2549
MemAE		0.7957	0.3863	0.2932	0.5593	0.1377	0.1039	0.5594	0.2173	0.2124
AnoDDPM		0.7850	0.4319	0.3259	0.7842	0.1765	0.1142	0.7842	0.3131	0.2075
AnoPILaD		0.8097	0.4322	0.3311	0.8312	0.3098	0.2326	0.8312	0.5420	0.4275

At the WSI-level, we further confirmed our findings from patch-level anomaly detection. Table 3 shows the WSI-level anomaly detection results for the four top-performing models (AE, MemAE, AnoDDPM, and AnoPILaD) in patch-level anomaly detection. For the classification of normal and abnormal slides, AnoPILaD was superior to all competing models across both datasets, evaluation metrics, and scoring strategies (Z_{MAX} and Z_{99}). The performance of AnoPILaD and other three models substantially varied between LH and C16. In LH, the four models obtained 0.7840 \sim 0.9943 AUC and 0.6616 \sim 0.9948 AUPR, whereas their performance dropped in C16, with 0.4551 \sim 0.6745 AUC and 0.3885 \sim 0.6140 AUPR. Nonetheless, AnoPILaD exhibited the smallest best performance drop of 0.3092 \sim 0.3576 AUC and 0.3600 \sim 0.4046 AUPR, highlight its robustness across datasets and organ types. In regard to large metastatic regions (C16 Macro), the performance generally declined for all models; however, AnoPILaD obtained

AUCs ranging from 0.8023 to 0.8062, suggesting its superior potential for cross-organ anomaly detection in lymph nodes.

Evaluating the segmentation results of metastatic regions, the strength of AnoPILaD was obvious, achieving the highest scores for all evaluation scenarios. We also observed that there are clear differences between the two anomaly detection approaches. Although AE-based methods (AE and MemAE) obtained comparable classification performance with diffusion-based methods (AnoPILaD and AnoDDPM), their segmentation performance was substantially poorer. As shown in Fig 3, MemAE assigns similar scores for all pixels in a slide, indicating a lack of specificity in detecting metastatic regions. The behavior of AE is almost identical to MemAE. These results suggest the importance of segmentation performance in the evaluation of anomaly detection, as classification alone may not accurately assess a model’s ability to localize abnormal regions.

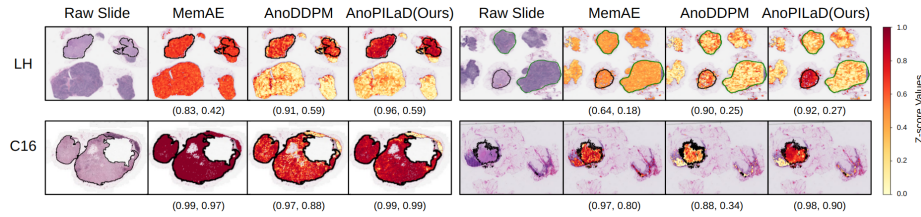


Fig. 3. Z-score heatmaps for four metastasis slides from two datasets with black contour showing metastasis annotation and green contour showing normal annotations. Numbers in parentheses denotes (Z_{99} , Dice) per slide.

We visualize and compare the reconstruction results by AnoPILaD and AnoDDPM (Fig 1). Starting from raw images, AnoPILaD and AnoDDPM aim to reconstruct normal-like structures with and without textual guidance, respectively. For an in-distribution sample (first row), both models successfully generate images with normal histologic features, preserving small, dense lymphocytic structures. However, as for OOD samples (second and third rows), which contain metastatic regions with disrupted tissue architecture, the differences between two models become more apparent. AnoDDPM partially reconstructs metastatic regions and fails to fully suppress pleomorphic nuclei and fibrotic tissue, leading to residual abnormalities that obscure the boundary between normal and abnormal structures. In contrast, with the guidance of text prompts, AnoPILaD generates images with more uniform lymphocytic arrangements, while suppressing distortions in tissue architecture. This results in a clearer distinction between normal and metastatic regions.

5 Conclusion

We propose AnoPILaD, a pathology-informed LDM for unsupervised anomaly detection in lymph nodes. By leveraging histological context provided through prompts, AnoPILaD introduces a stronger inductive bias during the reconstruction process, enhancing sensitivity to abnormal features and improving detection performance. Evaluating both patch-level and slide-level performance across two organ types, AnoPILaD substantially outperforms other anomaly detection methods including both density- and reconstruction-based approaches. The future study will entail the extension of AnoPILaD to further enhance its performance on cross-organ datasets, improving its adaptability and robustness in broader pathological applications.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant (No. RS-2025-0058322 and RS-2024-00397293) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (No. P0022543).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Nathanson, S. David. "Insights into the mechanisms of lymph node metastasis." *Cancer* 98.2 (2003): 413-423.
2. Budginaite, Elzbieta, et al. "Computational methods for metastasis detection in lymph nodes and characterization of the metastasis-free lymph node microarchitecture: A systematic-narrative hybrid review." *Journal of Pathology Informatics* 15 (2024): 100367.
3. Salehi, Mohammadreza, et al. "A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges." arXiv preprint arXiv:2110.14051 (2021).
4. Kingma, Diederik P. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).
5. Xiao, Zhisheng, Qing Yan, and Yali Amit. "Likelihood regret: An out-of-distribution detection score for variational auto-encoder." *Advances in neural information processing systems* 33 (2020): 20685-20696.
6. Ren, Jie, et al. "Likelihood ratios for out-of-distribution detection." *Advances in neural information processing systems* 32 (2019).
7. Serrà, Joan, et al. "Input complexity and out-of-distribution detection with likelihood-based generative models." *arXiv preprint arXiv:1909.11480* (2019).
8. Hinton, Geoffrey E., and Richard Zemel. "Autoencoders, minimum description length and Helmholtz free energy." *Advances in neural information processing systems* 6 (1993).
9. Chong, Yong Shean, and Yong Haur Tay. "Abnormal event detection in videos using spatiotemporal autoencoder." *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21-26, 2017, Proceedings, Part II* 14. Springer International Publishing, 2017.

10. Gong, Dong, et al. "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
11. Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.
12. Schlegl, Thomas, et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." *International conference on information processing in medical imaging*. Cham: Springer International Publishing, 2017.
13. Schlegl, Thomas, et al. "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks." *Medical image analysis* 54 (2019): 30-44.
14. Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.
15. Wyatt, Julian, et al. "Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
16. Graham, Mark S., et al. "Denoising diffusion models for out-of-distribution detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.
17. Linmans, Jasper, et al. "Diffusion models for out-of-distribution detection in digital pathology." *Medical Image Analysis* 93 (2024): 103088.
18. Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.
19. Zhang, Jingyi, et al. "Vision-language models for vision tasks: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).
20. Lu, Ming Y., et al. "Towards a visual-language foundation model for computational pathology." *arXiv preprint arXiv:2307.12914* (2023).
21. Bejnordi, Babak Ehteshami, et al. "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer." *Jama* 318.22 (2017): 2199-2210.
22. Heusel, Martin, et al. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." *Advances in neural information processing systems* 30 (2017).
23. Diffusers: State-of-the-art diffusion models" by Patrick von Platen, Suraj Patil, Anton Lozhkov, et al. (2022), available on GitHub as a repository named "hugging-face/diffusers
24. Liu, Luping, et al. "Pseudo numerical methods for diffusion models on manifolds." *arXiv preprint arXiv:2202.09778* (2022).
25. Virtanen, Pauli, et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature methods* 17.3 (2020): 261-272.